ChinaXiv合作期刊

宽视场成像网格化算法中 w-plane 最优经验值研究

于晓雨 [1] 邓辉 [1,2] 梅盈 [1] 卫守林 [1] 石聪明 [1] 王威 [3] 戴伟 [1] 王锋 [1,2]
（1. 昆明理工大学云南省计算机技术应用重点实验室；2. 广州大学天体物理中心；3. 中国科学院国家天文台）

摘要： 射电干涉阵列宽视场成像网格化处理过程中必须考虑 w 项的影响。w-projection 和 w-stacking 是两个重要的宽视场成像网格化处理算法， w-plane 参数是算法中影响计算速度和成图质量的一个重要因素。本文研究了 w-projection 和 w-stacking 两种网格化算法，利用 SKA1 低频阵台站数据和 ASKAP 软件包进行模拟观测，对两种算法在不同 w-plane 参数取值情况下的成图速度和成图质量进行了定量分析对比。结果进一步表明 w-plane 是性能改善的重要参数，针对 w-projection 算法来说，w-plane 取值应比一般给定的经验值要大才能较好的得到成像效果。w-stacking 算法虽然有很大的速度优势，但算法实现中 w-plane 的影响更为显著，论文给出了推荐的 w-plane 取值。本文工作是大视场成像算法的基础研究工作，对未来 SKA 科学数据处理中的管线设计有较好的参考价值。

关键词： SKA-SDP 宽视场成像 网格化算法
中图分类号： P161 **文献标识码：** A **文章编号：** 1672-7673(2018)xx-xxxx-xx

平方公里阵 SKA (square kilometer array) 将是国际上建造的最大综合孔径射电望远镜[1]。SKA 由数千个反射面天线和多达一百万个低频天线组成，总接收面积达到一平方公里。

SKA 科学数据处理(Science Data Processor, SDP)是 SKA 设计中的一个重要组成部分。SDP 专注于设计硬件平台、软件和算法来处理射电望远镜的观测数据[2]。SDP 的设计面临的挑战是多方面的。SDP 需要处理来自望远镜的海量观测数据量。SKA1(约占 10%的总体规模)数据量将达到 300PB/年。为了满足 SKA 科学研究开展的需要，SDP 在数据处理管线设计中，需要充分地考虑数据处理的性能与质量。由于 SDP 处理需要的计算力高出现有最大的天文数据处理系统两个数量级，SDP 的关键算法的运行速率要达到百万兆级的运算水平。SKA2 的计算力需求估计需要达到 4 Exaflops 的水平。因此，开展对 SDP 数据处理中关键算法的研究，提高数据处理的性能具有显著的意义。

网格化（Gridding）算法是综合孔径望远镜数据处理中的一个重要算法，对计算的消耗占据了很大比重。对网格化算法进行性能优化，对 SDP 有显著的价值和意义。低频射电望远镜大视场成像受到很多条件的制约，其中最重要的一个影响因素是非共面基线效应。在大视场与非共面基线效应下，对可见度数据直接进行傅立叶变换会导致图像严重畸变，因此，在大视场成图时，解决 w 项的问题至关重要。目前，应对大视场与非共面基线效应的算法有：faceting[3]、三维傅立叶变换[4]、w-projection[5]、w-stacking[6] 和 warped snapshots[4]等算法以及其他复合算法。在这些方法中，w-projection 和 w-stacking 是目前应用相对较广的网格化算法。其中 w-projection 算法在牺牲较大内存的情况下，具有优越的计算速度和误差控制。该方法不仅可以使用传统的洁化(CLEAN)方法去卷积，还可以选择多尺度洁化方法，在高动态范围下，有较大的优势，相位中心基本不失真。w-stacking 算法在图像空间以消耗更多的内存为代价处理傅里叶空间中的 w 项，计算速度优于 w-projection 快[7]。

在 w-projection 和 w-stacking 算法的具体实现过程中，w-plane 是一个重要参数。在 w 的最大值和最小值之间分出多个 w-plane，对不同 w-plane 的可见度数据使用不同的卷积核进行卷积运算。因此，w-plane 的数量直接影响到算法的计算速度、内存开销以及成图质量。本文主要对 w-projection 和 w-stacking 算法在实现与实际运行中的 w-plane 进行实验研究，利用 SKA1 低频阵台站数据和 ASKAP 软件包进行模拟观测，以期掌握 w-plane 对整体计算性能的影响，为后期的 SKA-SDP 建设以及国内科学数据中心建设中的管线系统设计提供参考。

基金项目：国家重点研发计划(2018YFA0404603,2016YFE0100300)，国家自然科学基金委员会-中国科学院天文联合基金资助重点项目(U1831204)，国家自然科学基金委员会-中国科学院天文联合基金资助项目(U1531132,U1631129)，国家自然科学基金资助项目(No.11403009,11463003,11773012, 11773043)，广州大学“创新强校工程”项目(2017KZDXM062)，云南省应用基础研究项目(2017FB001)，赛尔网络下一代互联网技术创新项目(No.NGII20170204)资助.
作者简介：于晓雨，男，硕士，研究方向：计算机应用技术. Email:yuxiaoyucnlab@sina.com
通信作者：邓 辉，女，教授，研究方向：天文图像、科学数据处理. Email:denghui@astrolab.cn

chinaXiv:201812.00018v1

1、w-plane

低频射电望远镜大视场成像最重要的一个影响因素就是非共面基线效应，在非共面基线效应的影响下 w 项的值将远远大于 1，因此直接对可见度数据进行二维傅里叶变换的条件已经不满足，针对这一问题，天文学家提出了一系列算法用来应对非共面基线效应造成的图像畸变。w-projection 算法通过将不同 w 值的可见度数据投影到 w=0 平面，从而消除 w 项的影响，有效提高成图质量。w-stacking 算法能够在像平面矫正非共面基线效应引起的相位偏移从而有效还原天空亮度分布。

1.1 w-projection 算法中的 w-plane

在射电综合孔径成像处理过程中，可见度数据与天空亮度之间的关系如公式（1）所示：

$$V(u,v,w)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{I(l,m)}{\sqrt{1-l^2-m^2}}e^{-2\pi i[ul+vm+w\sqrt{1-l^2-m^2}-1]}dldm \tag{1}$$

(4)

w-projection 算法将不同 w 值的可见度数据投影到 w=0 平面，公式（1）简化为 (2):

$$V(u,v,w=0)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{I(l,m)}{\sqrt{1-l^2-m^2}}e^{[-2\pi i(ul+vm)]}dldm \tag{2}$$

经过简单的指数运算，式（1）变化为式（3）和（4）：

$$V(u,v,w)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\frac{I(l,m)}{\sqrt{1-l^2-m^2}}G(l,m,n)e^{-2\pi i[ul+vm]}dldm \tag{3}$$

$$G(l,m,n)=e^{-2\pi iw(\sqrt{1-l^2-m^2}-1)} \tag{4}$$

公式（3）、（4）可以改写为公式（5），其中卷积核通过公式（6）近似计算得到：

$$V(u,v,w)=\hat{G}(u,v,w)*V(u,v,w=0) \tag{5}$$

$$\hat{G}(u,v,w)=\frac{i}{w}e^{-\pi i[\frac{u^2+v^2}{w}]} \tag{6}$$

上述过程显示，在 w-projection 算法中，卷积核是算法的核心。理想情况下，位于不同 w-plane 的可见度数据，在计算过程中应该采用不同的卷积核。在实际计算过程中，一般计算 w 的最大和最小值，在最大和最小值之间平均分成多个 w-plane，然后对每个 w-pane 按照公式（6）计算卷积核并保存下来，最后对每个可见度数据与其距离最近的 w-plane 的卷积核进行卷积，从而实现到 w=0 平面的投影。

显而易见，w-plane 的值越大，投影效果越好，但是计算量和内存消耗大；w-plane 的值越小，计算量和内存消耗小，但是投影效果欠佳。当 w-plane 的数量为零，则直接退化为二维傅立叶变换。Cornwell 等在前期研究中取最大的 w 值的平方根作为经验值[5]。

2 w-stacking 算法中 w-plane

w-stacking 算法的提出也是为了消除非共面基线效应对大视场成像的影响，但是不同于 w-projection 算法，w-stacking 并没有消除 w 项，而是对其进行了修正。

根据公式(1)将 uv 空间转换到 lm 空间得到式（7）：

$$\frac{I(l,m)}{\sqrt{1-l^2-m^2}}=e^{2\pi iw(\sqrt{1-l^2-m^2}-1)}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}V(u,v,w)e^{2\pi i(ul+vm)}dudv \tag{7}$$

chinaXiv:201812.00018v1

对公式(7)的两边在 w 范围内进行积分得到式（8）：

$$\frac{I(l,m)(w_{\max}-w_{\min})}{\sqrt{1-l^2-m^2}}=\int_{w_{\min}}^{w_{\max}}e^{2\pi iw(\sqrt{1-l^2-m^2}-1)}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}V(u,v,w)e^{2\pi i(ul+vm)}dudvdw \tag{8}$$

其中 $w_{\max}$ 和 $w_{\min}$ 代表 w 的最大值和最小值。式(8)的等号右侧包括两部分，一是二维傅立叶变换，二是对 w 的复数积分。对式（8）的等号右侧离散化得到式（9）：

$$\frac{I(l,m)(w_{\max}-w_{\min})}{\sqrt{1-l^2-m^2}}=\sum_{a=w_{\min}}^{w_{\max}}e^{2\pi uw_a(\sqrt{1-l^2-m^2}-1)}\mathscr{F}^{-1}\{V(u,v,w_a)\} \tag{9}$$

其中 $\mathscr{F}^{-1}\{V(u,v,w_a)\}$表示将可见度数据在特定的离散 w-plane 上进行网格化。

w-stacking 算法的实现步骤如下：

(1) 根据可见度数据 w 的范围设定 w-plane 的细分粒度；
(2) 依据每个可见度数据的 w 值，将可见度数据网格化到某一 w-plane;
(3) 对每个 w-plane 的可见度数据进行二维傅立叶变换得到相应图像;
(4) 根据 w 值对各个 w-plane 的图像进行相位矫正;
(5) 加权叠加所有图像，得到最终图像。

在 w-stacking 算法实现过程中，w-plane 分的过粗会导致产生伪影，分的过细又会增大计算量和内存消耗。

同时，w-stacking 算法将不同 w-plane 上的可见度数据看成是相互独立，各个面的数据可以并行进行二维傅立叶变换和相位改正，因此，w-stacking 算法在并行实现方面具有优势。

3 观测模拟仿真实验

3.1 实验环境

为了验证 w-plane 的影响，以期获得较为理想的 w-plane 取值，在确保成像质量的情况下降低成像处理时间，本文进行了一系列实验。数据处理软件采用了澳大利亚 SKA 探路者(Australian SKA Pathfinder, ASKAP)的数据处理软件系统[8]。ASKAP 通过名为 ASKAPsoft 的软件系统进行望远镜观测控制和数据处理[9]，ASKAPsoft 框架由三部分组成：望远镜观测系统、处理中心和科学数据存档。ASKAP 通过 bash shell 脚本组合 ASKAPsoft 中的程序形成数据处理管线，主要的数据处理管线有数据摄取管线、定标管线和成图管线[10]。

本实验在 8 台汉柏服务器上完成，具体硬件配置如表 1 所示。

表 1 硬件设备参数

Table 1 The parameters of device

硬件设备	设备参数
服务器型号	C640 G3（8）
CPU	Intel Xeon E5-2620 v3
内存	128GB
网络	Infiniband + 10Gb 以太网

3.2 模拟观测与成像处理

为了分析 w-plane 对 w-projection 和 w-stacking 中的影响，我们采用了模拟观测、成像、对后续图像进行分析的方法来验证不同的 w-plane 造成的影响。

1、 模拟观测

实验采用 SKA1-Low 望远镜的 512 个天线台站坐标数据，天线布局如图 1 所示。天空模型由 73 个点源构成，整体形状为“米”字，如图 2 (a)所示。实验采用 16 个计算节点，每个计算节点处理一个通道的数据，通道带宽 1MHZ，因此观测频率覆盖范围 16MHz，积分时间间隔 150s，模拟观测时长 6 小时。

chinaXiv:201812.00018v1

SKA1-Low 天线频率覆盖范围 50-350MHz，本实验中心参考频率为 100MHz，设置 16 个带宽为 1MHz 的通道，因此观测频率范围为 92-108MHz。

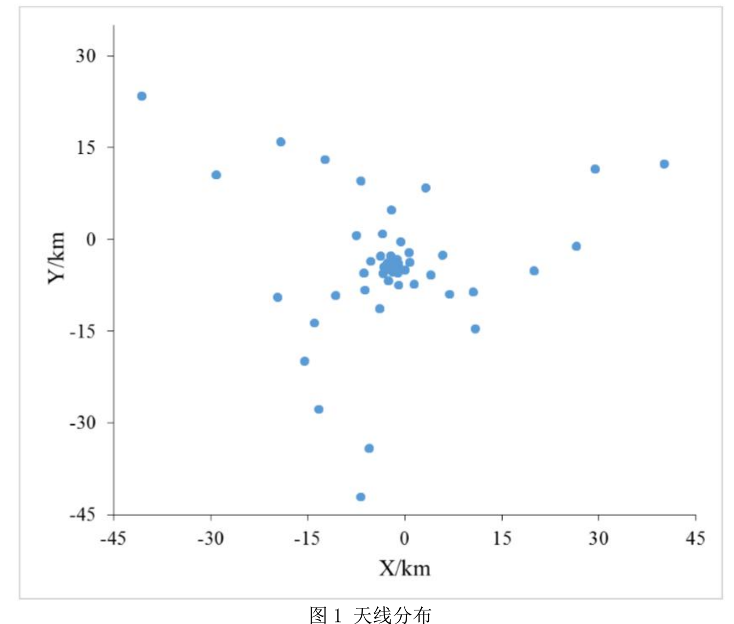


图 1 天线分布
Fig. 1 Antennas layout

使用模拟观测程序对天空模型进行 16 个通道的模拟观测，得到 18837504 行 5.5GB 的模拟可见度数据。

2、成像处理

使用 w-projection 和 w-stacking 算法，对模拟观测获得的可见度数据 16 个通道进行连续谱成图。ASKAP 成图管线在 8 个计算节点上并行运算，每个计算节点处理两个通道。实验过程中，不断改变 w-plane 的数量，得到不同 w-plane 设置下的图像，并记录每次成图时间供后续分析使用。图 4（a）显示了随着 w-plane 数量的增长，两种算法处理时间的变化。

图 2 (b)是二维傅立叶处理得到的图像，可以看到 w 项影响导致的拖尾效应(Smearing)非常明显。图 3 是 w-plane 数量设置为 171 的情况下，w-projection 和 w-stacking 算法处理得到的图像，拖尾效应得到抑制，但边缘的模拟星并没有被很好地还原。

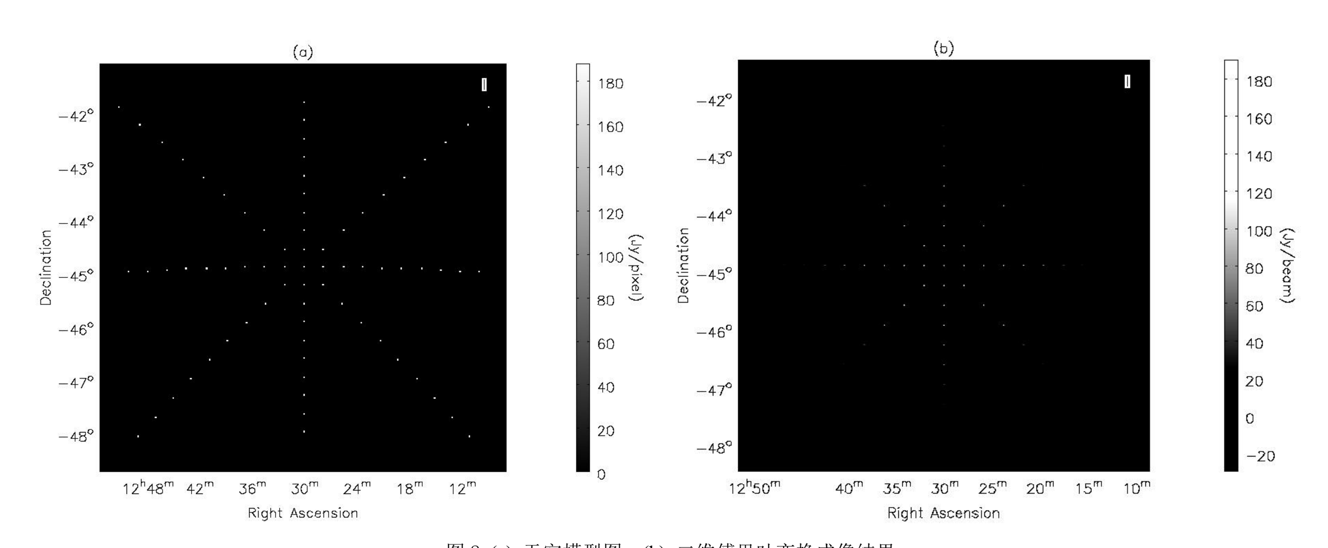


图 2 (a) 天空模型图；(b) 二维傅里叶变换成像结果
Fig. 2 (a) Sky model image; (b) The result of two-dimensional Fourier transform imaging

chinaXiv:201812.00018v1

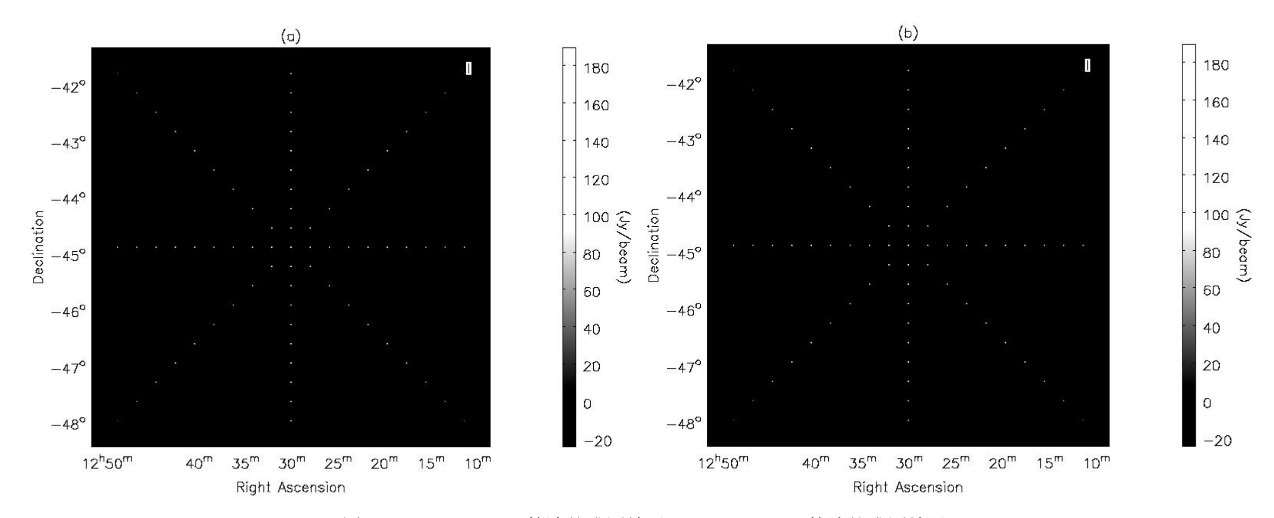


图 3 (a) w-projection 算法的成图结果；(b) w-stacking 算法的成图结果

Fig. 3 (a) The result of the w-projection imaging; (b) The result of the w-stack imaging

3、图像质量评价方法

为了评价不同 w-plane 数量下的两种算法的成像质量，研究采用了 SELAVY 软件包，该软件包是 Duchamp 星源搜索算法在 ASKAP 软件中的实现模块，用来搜索观测结果中的星数量。由于在仿真观测时已经预设了总的星数为 73，同时所有星的位置已知。因此，通过 SELAVY 搜索出的星与已知天空模型星进行对比，利用正确还原的星数量来快速判断成像的质量。图 4（b）显示了随着 w-plane 数量的增长，SELAVY 搜索出点源数量的变化。

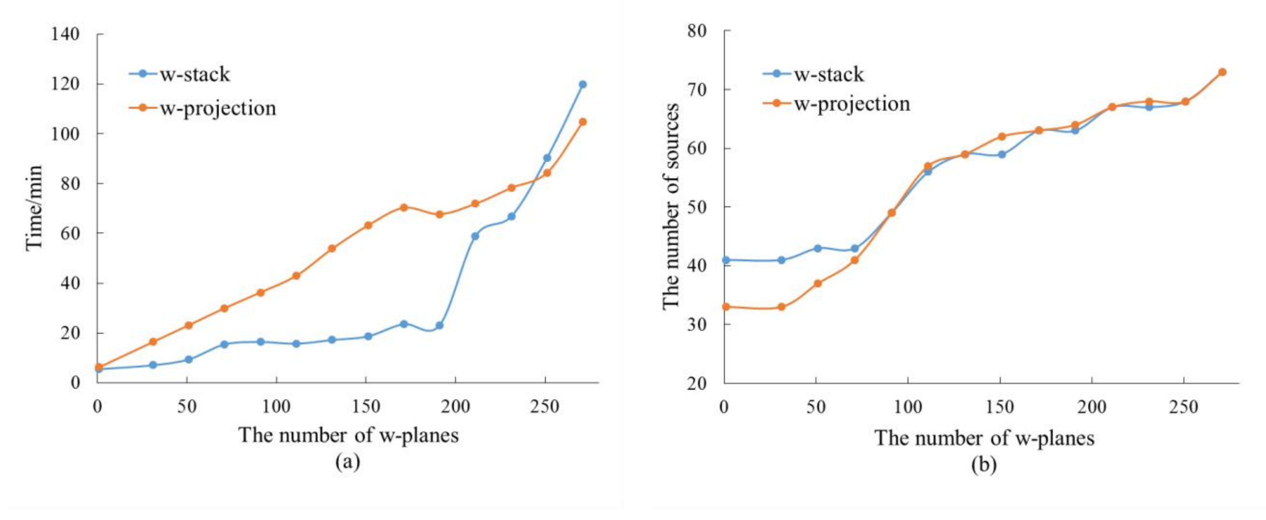


图 4 (a)计算时间与 w-plane 的关系；(b) SELAVY 搜索出点源数量与 w-plane 的关系

Fig. 4 (a) The relation between the running time and w-plane; (b) The relation between the number of searched sources and w-plane

3、结果分析

本实验最大基线长度 81602 米，最大 w 约为 28000 个波长。根据 Cornwell 等前期研究的结果，w-plane 值一般设置为最大 w 的平方根，也就是 167 。

从图 4 (a)可以看到 w-projection 的运行时间随着 w-plane 值的增加呈近似线性增长。而 w-stacking 算法的运行时间则表现为：当 w-plane 的值小于 $\sqrt{w}$ 的值的情况下其计算时间明显小于 w-projection 算法；w-plane 的值大于 $\sqrt{w}$ 的值后其增长速度很快并在 w-plane 大于 1.5 $\sqrt{w}$ 之后超过 w-projection 算法的运算时间。这是由于大图像成像时太多的 w-plane 需要进行相位的改正从而导致 w-stacking 算法的计算力消耗大增[11]，w-stacking 算法也就失去了速度优势。

SELAVY 软件包对两种算法在不同的 w-plane 下成图结果进行星像统计。从图 4 (b)可以看到在 w-plane 较小的情况下 w-stacking 算法成图得到星的数量多于 w-projection 的结果，但是随着 w-plane 的增加两者的成图质量近似相同。

chinaXiv:201812.00018v1

ChinaXiv合作期刊

w-plane 的取值不同于以往的经验值 $\sqrt{w}$ ，实验可以看出：在 w-plane 大于 1.5 $\sqrt{w}$ 之后成图质量更好，SELAVY 程序包可以得到所有的 73 个星源的数据，因此为获得高分辨率的天空图 w-plane 的取值应大于 1.5 $\sqrt{w}$ 。虽然两者的成图质量相同但是 w-stacking 算法在 w-plane 的值小于 $\sqrt{w}$ 的值的情况下相较于 w-projection 具有很大的速度优势，但当 w-plane 的值大于 $\sqrt{w}$ 后，w-stacking 的速度优势将不存在。

通过分析这两种算法的成图质量和运行时间趋势，可以看到 w-stacking 算法更适合 w-plane 值小于 $\sqrt{w}$ 的情况，为弥补其 w-plane 较小引起的成像准确度的问题，今后可以考虑通过结合多波束观测进行成图，再通过 SKA 的 LINMOS(Linear Mosaic Applicator)软件包实现多波束成图后的拼接，该软件包实现一系列图片的线性拼接，从而提高成像质量。w-projection 算法则更适合通过提高 w-plane 的值实现更为准确的成图。

4. 结论

综上所述，针对 w-projection 算法和 w-stacking 算法在管线中的应用，有如下结论：

1、合理的 w-plane 的取值对于成像有关键的作用。实验表明，w-plane 的取值应是经验取值 $\sqrt{w}$ 的 1.5 倍，此时 w-projection 算法和 w-stacking 算法都可以得到很好的成图质量，这与 Cornwell 估计的有一定的偏差。

2、w-plane 的取值对于结果的正确性有紧密的关系。事实上只有 w-plane 取值较大时才有可能完整地恢复出天空信息。未来在数据处理时，应在准确度要求和计算时间需求方面进行折衷。当 w-plane 的值小于 $\sqrt{w}$ 的时候，w-stacking 算法比 w-projection 算法有很大的计算时间优势。对于成图分辨率要求高且 w-plane 的值大于 $\sqrt{w}$ 的时候，w-stacking 算法将失去速度优势。

本文的工作表明 SKA-SDP 数据处理管线应根据不同的需求选择不同的 gridding 算法，后续工作可以在运算速度提高，提高算法的并行性为研究重点，以期获得更好的处理能力。

参考文献

[1] Dewdney P E, Hall P J, Schilizzi R T, et al. The square kilometre array[J]. Proc IEEE, 2009, 97: 1482–1496
[2] P. Alexander, V. Allan, et al. PDR.01 SDP Architecture[EB/OL].[2018-09-13]. http://www.astron.nl/~broekema/papers/SDP-PDR/PDR01%20System%20Architecture.pdf
[3] Cornwell T J, Perley R A. Radio-interferometric imaging of very large fields - The problem of non-coplanar arrays[J]. Astronomy & Astrophysics, 1992, 261(1):353-364.
[4] Perley R A. Synthesis Imaging in Radio Astronomy II[C].San Framcisco:NRAO/NMIMT,1999,180:383
[5] Cornwell T J, Golap K, Bhatnagar S. The Noncoplanar Baselines Effect in Radio Interferometry: The W-Projection Algorithm[J]. IEEE Journal of Selected Topics in Signal Processing, 2008, 2(5):647-657.
[6] B. Humphreys B, Cornwell T, Humphreys B, et al. Analysis of Convolutional Resampling Algorithm Performance[EB/OL].[2018-09-13]. www.skatelescope.org/pages/page_memos.html
[7] 劳保强，安涛，陈骁,等. 低频射电望远镜阵列宽视场成像技术研究[J]. 天文学报, 2017, 58(5):108-129.
Lao Baoqiang, An Tao, Chen Xiao, et al. Research on Wide-field Imaging Technologies for Low-frequency Radio Array[J]. Acta Astronomica Sinica, 2017, 58(5):108-129.
[8] Deboer D R, Gough R G, Bunton J D, et al. Australian SKA Pathfinder: A High-Dynamic Range Wide-Field of View Survey Telescope[J]. Proceedings of the IEEE, 2009, 97(8):1507-1521.
[9] Tim Cornwell, Ben Humphreys, et al. ASKAP Science Processing[EB/OL]. [2018-09-13]. http://www.atnf.csiro.au/projects/askap/ASKAP-SW-0020.pdf
[10] Chapman J M. CASDA: The CSIRO ASKAP Science Data Archive[J]. Iau General Assembly, 2015, 22.
[11] Goodrick L. Image Reconstruction in Radio Astronomy with Non-Coplanar Synthesis Arrays[D]. Matieland: Electrical and Electronic Engineering University of Stellenbosch, 2015.

A Study on Optimal Experiential W-plane for Wide-Field Gridding Algorithm

Yu Xiaoyu[1], Deng Hui[1,2], Mei Ying[1], Wei Shoulin[1], Shi Congming[1], Wang Feng[1,2]
(1. Key Laboratory of Computer Technology Application; 2. Center for Astrophysics, Guangzhou University)

Abstract: Astronomy data processing platform is facing unprecedented big data challenge. SKA(square kilometer array) as the biggest radio synthesis telescope, its generated data rate of averaged visibilities will be hundred gigabit per second. SKA-SDP handles big data and extremely large computing cost by data processing pipeline. Gridding algorithm will take a large proportion of compute load. How to save power cost and improve the computational efficiency, on the prerequisite of guaranteed imaging quality, need to be well-considered. W-

chinaXiv:201812.00018v1

projection and w-stacking are two gridding algorithms possibly be used in SKA-SDP; The present paper study on the effect of w-plane in those algorithms. Based on SKA continuum imaging pipeline and simulation observation, we analyze the effect of w-plane to the imaging results. Our experimental results show that w-plane is a crucial parameter for performance promotion, and w-plane should be bigger than experimental value in w-projection algorithm. Compare with w-projection algorithm, w-stacking have advantage in run time, yet w-plane have great effect on w-stacking algorithm. The present paper is a basic work for study on large field imaging, and a new w-plane for w-projection and w-stacking algorithm is proposed. Our study can serve as a reference for designs of SKA data processing pipeline.

Key words: SKA-SDP; Wide-field imaging; Gridding algorithm

chinaXiv:201812.00018v1